



ANÁLISIS DE LAS SERIES SINTÉTICAS DE KT PARA UNA LOCALIDAD, GENERADAS CON DISTINTAS BIBLIOTECAS DE MATRICES DE TRANSICIÓN DE MARKOV.

Rosana Aristegui, Raúl Righini, Federico Taddei

GERSolar, INEDES, Universidad Nacional de Luján, Av. Constitución y Ruta 5, (6700) Luján,
Buenos Aires, Argentina – Tel. (54-02323)-440241– e-mail: gersolar@yahoo.com.ar

RESUMEN: Se generan series sintéticas que brindan un año de valores diarios de índice de claridad (kt) para la localidad de Luján (Prov. de Bs. As.), utilizando valores medios mensuales históricos y cuatro bibliotecas de 10 matrices de transición de Markov. Estas matrices contienen la información de las probabilidades de transición de los valores diarios de una clase de kt a otra, en días consecutivos, de cuatro distintas bases de datos. Tres de estas bibliotecas fueron construidas dentro de este trabajo con distintas bases de datos de estaciones argentinas. La cuarta biblioteca corresponde a una base de localidades en el hemisferio norte y fue obtenida de la bibliografía. Se comparan estimadores estadísticos para las series generadas con las cuatro bibliotecas de matrices con el fin de evaluar cuáles de las bibliotecas generan las mejores series de datos de kt.

PALABRAS CLAVE: Series sintéticas, Matrices de transición de Markov, índice de claridad.

INTRODUCCIÓN

Tal como se mencionara en un trabajo anterior (Aristegui et al., 2010) para realizar estudios detallados de diseño y predicción de la respuesta de un sistema aplicado de aprovechamiento de la energía solar, se requiere contar con series continuas de distinta extensión temporal de valores diarios de irradiación solar incidente en el sitio de instalación. Usualmente los valores medios mensuales de la irradiación solar global al nivel de la superficie terrestre son medidos o estimados usando modelos que emplean datos satelitales, mediciones de heliofania o algoritmos asociados a redes neuronales. Debido a que los modelos de simulación suelen requerir de datos diarios de kt es importante contar con series sintéticas cuyos valores diarios estimados sean estadísticamente comparables a las series reales.

Existen varias maneras de generar estas series sintéticas. Algunos métodos son auto regresivos, mientras que otros tienen más éxito usando análisis markoviano a primer orden (A.Maafi et al., 1998, A.Sangisolo, 1997; Kamal et al., 1999).

Usando las Matrices de Transición de Markov (MTM) se pueden generar series estadísticamente significativas que sean comparables a las series reales, tal como comprobó Aguiar (1988). Específicamente Frulla (Frulla et al., 1991) y Mellit (Mellit et al. 2005) las emplearon para estimar datos diarios de radiación usando modelos basados en imágenes satelitales y redes neuronales.

El método MTM emplea matrices construidas en base a las probabilidades de transición de los valores de kt diarios. Esos valores pueden agruparse en 10 clases diferentes, por lo que en el actual trabajo empleamos 10 de ellas. Los presupuestos estadísticos que justifican el uso del método son dos:

1. Cualquier valor de radiación diaria muestra una significativa correlación sólo con el valor de radiación del día precedente. (Amato et al. 1987).
2. La función densidad de probabilidad de los valores diarios de kt es la misma para todos los meses cuyo índice de claridad medio pertenezca a la misma clase.

Las matrices así obtenidas pueden ser usadas para generar secuencias de kt diarios para cualquier localidad, conociendo sólo los valores medios mensuales históricos <kt> en ese lugar. De acuerdo al valor medio mensual histórico de un dado mes se selecciona la matriz de Markov que se usará y se genera con ella una serie de valores diarios.

MATERIALES Y MÉTODOS

Para generar las tres matrices de transición de Markov con información de Argentina se emplearon datos provenientes de dos fuentes:

- a) La Red Solarimétrica Argentina.
- b) La estación de medición de la radiación solar situada en Luján.

Los datos seleccionados de la Red Solarimétrica pertenecen a series históricas de datos de 28 estaciones entre 1978 y 2008. Los datos de Luján abarcan el período 2006-2010.

Para generar las bibliotecas de matrices se seleccionaron, dentro de las bases de datos usadas, tres áreas:

Biblioteca 1) El área completa conformada por las 28 estaciones pertenecientes a la Red.

Biblioteca 2) Una subregión de la anterior, que comprende 10 estaciones pertenecientes a la zona fitogeográfica similar a la de la estación Luján.

Biblioteca 3) La estación Luján.

Abra Pampa	La Consulta	Rio Cuarto *
Alto Valle	La Rioja	Roque Sáenz Peña
Anguil *	Luján *	San Carlos
Balcarce *	Marcos Juárez *	San Juan
Cerrillos	Mercedes	San Miguel *
Cerro Azul	Oliveros *	Trelew
Colonia Santa Rosa	Paraná *	Ushuaia
El Colorado	Pergamino *	Villa Mercedes *
El Sombrerito	Rafaela *	
Famaillá	Rama Caída	

Tabla 1: Serie de estaciones usadas para la construcción de las matrices de Markov. Las indicadas con () corresponden al subconjunto de 10 estaciones pertenecientes a la misma zona geográfica que Luján.*

Se escribió un programa en Visual Basic que generaba las matrices y las series sintéticas de kt. El programa respetaba los procedimientos que requiere el proceso, los cuales se encuentran descriptos en Aguiar (1988). No se consideraron meses a los que faltaban más de cuatro días de datos para el cálculo de las probabilidades de transición. Los kt mayores a 0,9 no fueron considerados por estimarse que se trataba de información errónea. Para la generación de las series sintéticas se usó un proceso aleatorio iterativo, que se detenía cuando los valores medios mensuales de kt simulados diferían de las medias históricas en menos de 0,01.

Usando las tres bibliotecas de matrices de Markov (bibliotecas 1, 2 y 3) y la biblioteca Aguiar (biblioteca 4) se generaron con cada una de ellas cinco series anuales de valores diarios de kt para la ciudad de Luján, empleando los valores de <kt> históricos de esta estación.

Las series obtenidas fueron comparadas entre sí y con los valores diarios medidos en Luján.

RESULTADOS

Comparación de bibliotecas de Markov

Se compararon de a pares las bibliotecas de Markov empleadas. Dichas bibliotecas correspondían a las que tienen matrices de transición de Markov calculadas usando las bases de datos seleccionadas, y las que fueron obtenidas por Aguiar. Para hacerlo se calculó la suma de los cuadrados de las diferencias de sus elementos para cada una de las diez matrices de Markov correspondientes a cada clase de kt. A su vez esos diez valores fueron sumados entre sí para obtener un índice de diferenciación que represente las diferencias entre bibliotecas. Un índice más alto significa una diferencia mayor entre los pares de bibliotecas comparadas. En la tabla siguiente se muestran los resultados obtenidos.

<i>Pares de bibliotecas</i>	<i>Índice de diferenciación</i>
1-3	1952
4-1	1066
4-3	1806
1-2	1307
4-2	1502
3-2	1935

Tabla 2: Índice de diferenciación entre los pares de bibliotecas de Markov usadas: 1) Base de datos de 28 estaciones. 2) Base de datos de 10 estaciones. 3) Base de datos de Luján. 4) Base de datos de Aguiar.

Los índices de diferenciación permiten establecer que las bibliotecas de Aguiar y la obtenida usando la base de datos extensa que comprendía 28 estaciones de la Red Solarimétrica (1) son las que presentan mayor similitud. La biblioteca de matrices de Luján (3) resultó bastante diferente de todas las otras, lo que se ve reflejado en que los valores de sus índices de diferenciación con el resto de las bibliotecas son los más altos. Una conclusión evidente que se desprende es que las matrices de Luján son claramente diferentes de todas las obtenidas con bases de datos que no son tan fuertemente locales.

Comparación de series

Para estudiar la bondad de las series sintéticas de datos generadas, se compararon distintos estadísticos, que fueron calculados tanto para ellas como para la serie de datos medidos en Luján.

Autocorrelación

Se calcularon los coeficientes de autocorrelación con corrimientos de 1 a 10. Los resultados se muestran en la figura 1.

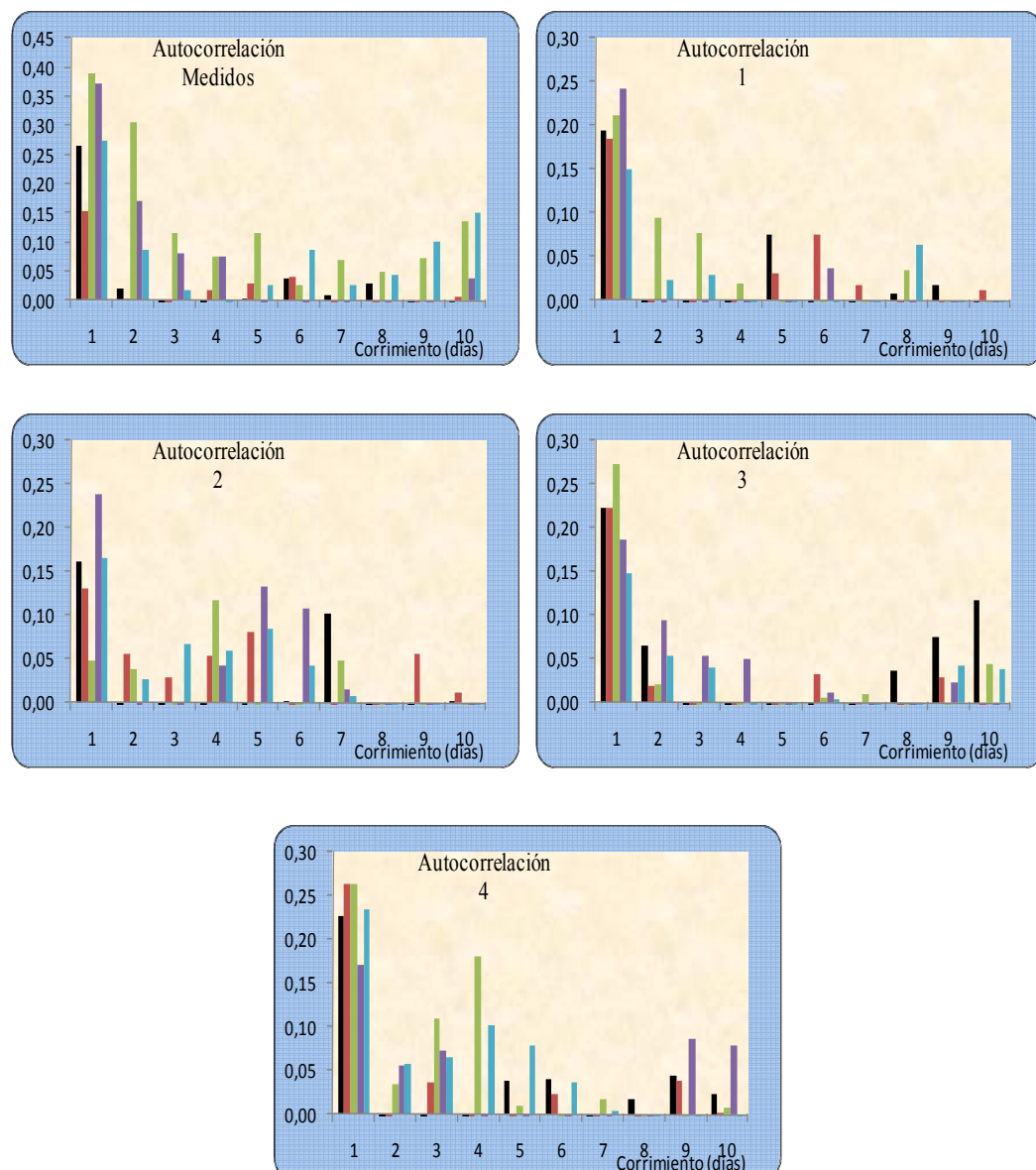


Figura 1: Coeficientes de autocorrelación correspondientes a las cinco series de años generados con las cuatro bibliotecas de Markov y la propia serie de datos medidos en Luján.

Los coeficientes muestran que, tal como era esperable, existe una fuerte dependencia entre la radiación de un día y el siguiente, disminuyendo a medida que los días se distancian. No es posible observar aquí diferencias sustanciales entre los resultados obtenidos con las distintas bibliotecas.

Desvío estándar

Se calcularon los desvíos estándar mensuales para todos las series generadas. Estos valores se compararon con los desvíos estándar mensuales promedio en los 5 años medidos mediante el cálculo de las diferencias relativas. En la generación de las series sintéticas no se impusieron restricciones en cuanto a la desviación estándar. A pesar de ello, los valores generados muestran un comportamiento aceptable. En términos generales las series generadas con las matrices de Aguiar fueron las de peor desempeño.

Mes	4-Med	1-Med	2-Med	3-Med
Enero	3,0	4,4	13,1	17,8
Febrero	29,1	8,5	13,9	4,0
Marzo	24,8	9,2	0,1	1,0
Abril	6,7	6,8	4,4	8,1
Mayo	1,0	5,2	10,9	15,3
Junio	4,0	2,0	4,9	7,5
Julio	17,7	8,0	18,5	10,5
Agosto	27,4	10,3	1,5	3,7
Septiembre	17,6	14,4	6,4	7,3
Octubre	20,1	10,7	14,3	20,5
Noviembre	13,0	18,8	1,4	17,8
Diciembre	13,9	9,3	7,4	4,2
Suma	178,2	107,6	96,9	117,9

Tabla 3: Diferencias relativas entre los desvíos estándar mensuales de las series sintéticas generadas con los datos medidos.

Días nublados consecutivos

Se analizó la ocurrencia de días nublados consecutivos totales para los cinco años sintéticos generados con las diferentes bibliotecas, y se las comparó con los datos medidos en la estación Luján. Los resultados se muestran en el gráfico y la tabla siguientes.

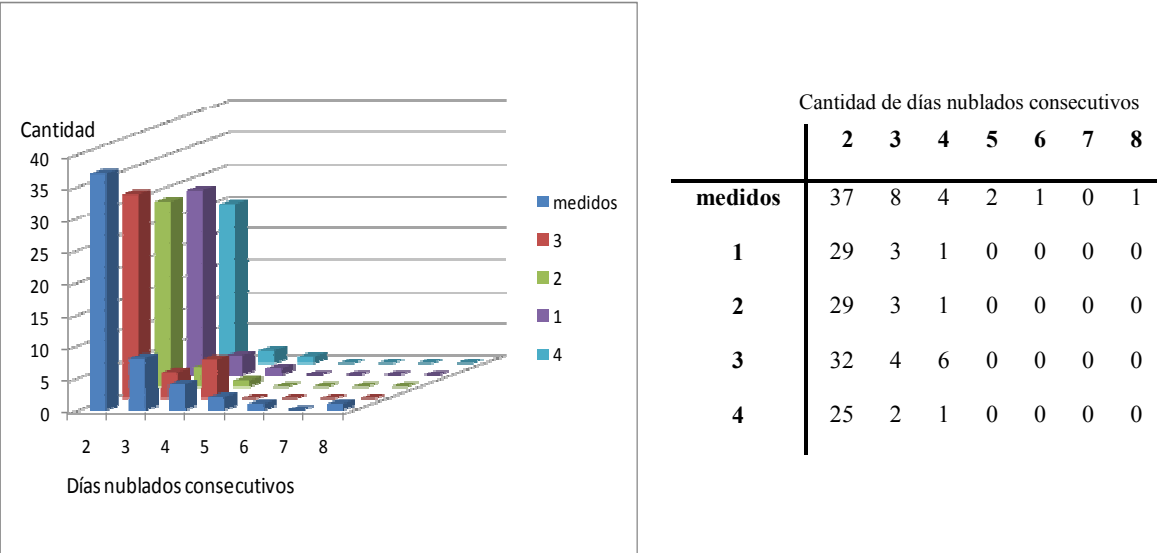


Figura 2 y Tabla 4: Frecuencia de días nublados consecutivos (totales en 5 años) para las series sintéticas generadas y los valores medidos.

Es posible ver aquí que, las simulaciones son todas razonables pero la creada con la serie de datos de Luján se ajusta mejor a la serie de datos reales.

Test de Smirnov

Usando el estadístico de Smirnov se compararon todos los años medidos entre sí y con las series sintéticas generadas con dichos años. El objetivo de este estadístico es comparar las funciones de distribución de probabilidad correspondiente a los datos medidos en la estación Luján con las de las series sintéticas. En la tabla siguiente se muestran los resultados.

Serie	2006	2007	2008	2009	2010	promedio
Medidos 2006	0,00	0,05	0,03	0,09	0,16	0,03
Medidos 2007	0,05	0,00	0,07	0,11	0,16	0,05
Medidos 2008	0,03	0,07	0,00	0,10	0,18	0,05
Medidos 2009	0,09	0,11	0,10	0,00	0,12	0,05
Medidos 2010	0,16	0,16	0,18	0,12	0,00	0,10
1	0,09	0,12	0,09	0,19	0,25	0,11
1	0,11	0,13	0,10	0,20	0,25	0,12
1	0,11	0,13	0,10	0,20	0,23	0,12
1	0,13	0,15	0,12	0,22	0,23	0,14
1	0,11	0,13	0,11	0,21	0,24	0,13
2	0,07	0,10	0,06	0,16	0,22	0,09
2	0,10	0,13	0,09	0,19	0,21	0,11
2	0,10	0,12	0,10	0,20	0,23	0,11
2	0,10	0,12	0,09	0,19	0,20	0,12
2	0,07	0,10	0,07	0,17	0,19	0,09
3	0,07	0,08	0,06	0,12	0,14	0,06
3	0,07	0,08	0,09	0,10	0,09	0,03
3	0,06	0,07	0,07	0,08	0,11	0,02
3	0,05	0,08	0,06	0,12	0,12	0,05
3	0,06	0,08	0,07	0,11	0,11	0,05
4	0,09	0,12	0,06	0,15	0,15	0,08
4	0,08	0,12	0,07	0,16	0,15	0,09
4	0,08	0,12	0,07	0,17	0,17	0,10
4	0,10	0,13	0,07	0,14	0,14	0,07
4	0,08	0,12	0,05	0,10	0,16	0,06

Tabla 34: Resultados del test de Smirnov para las series generadas y los datos medidos.

El estimador del test de Smirnov cuantifica el error cometido al rechazar la serie de datos como perteneciente a la misma distribución estadística tomada como referencia. Los valores más bajos indican mayor similitud entre las distribuciones de probabilidad. Es importante esta cuantificación ya que permite evaluar el desempeño de las series generadas por medio de un estimador generado de una forma no arbitraria.

De la comparación de los estimadores surge a la vista que las series generadas con los datos de la estación Luján son las que mejores describen los datos medidos. La conveniencia de usar una base local para simular años típicos queda así manifiestamente por encima de otras series generadas con matrices de Markov que no tengan en cuenta exclusivamente esos datos.

CONCLUSIONES

En primer lugar es evidente, en contra de lo afirmado por Aguiar (Aguiar, 1988), que las matrices que forman la biblioteca de matrices de transición del método de Markov dependen fuertemente de la cobertura geográfica de la base de datos. Los índices de diferenciación calculados así lo demuestran. Obviamente cuanto mayor sea la cobertura geográfica más similares resultan las bibliotecas entre sí. Sin embargo parece mejor usar una base de datos local para simular apropiadamente años sintéticos, tal como demuestran los estadísticos relacionados con la cantidad de días nublados consecutivos y el test de Smirnov. En particular la cantidad de días nublados consecutivos reviste gran importancia en la simulación del desempeño de equipos relacionados con el aprovechamiento energético de la irradiación solar.

El resto de los estadísticos analizados no muestran mayor diferencia entre los resultados brindados por las diferentes bibliotecas. Un trabajo futuro debería considerar una serie más extensa de datos, y diferentes áreas de cobertura de los mismos para arribar a conclusiones definitivas.

REFERENCIAS

- Aristegui R., Raichijk C., (2010) Generación de series sintéticas de valores diarios de irradiación solar usando matrices de transición de Markov desarrolladas a partir de datos de Argentina. *Avances en Energías Renovables y Medio Ambiente*, vol. 14, pp. 11.69-11.76, versión CD-ROM (2010).
- Maafi A., Adane A. (1998) Analysis of the performances of the first-order two-state Markov model using solar radiation properties. *Renewable Energy*, 13, 2, 175-193.
- Sansigolo A. (1997) Non-Stationary Markov Chains for Modelling Daily Sunshine at Silo Paulo, Brazil. *Theor. Appl. Climatol.* 56, 225-230
- Kamal L., Zahra Jafri Y. (1999) Stochastic modeling and generation of synthetic sequences of hourly global solar irradiation at Quetta, Pakistan. *Renewable Energy* 18, 565-572.
- Aguiar R.J., Collares-Pereira M., Conde J.P. (1988) Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices. *Solar Energy*, 40, 3, 269-279.
- Frulla L., Pettigiani E., Grossi Gallegos H., Atienza G. (1991) Construcción y análisis de series temporales de radiación solar global obtenidas a partir de observaciones GOES. *Memorias del V Simposio Latinoamericano de Percepción Remota*, Cusco, Perú, 188-202.
- Mellit A., Benghanem M., Hadj Arab A., Guessoum A. (2005) A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach. *Solar Energy*, 79, 469-482.
- Amato U., Andreta A., Bartoli B., Coluzzi B., Cuomo V., Fontana F., Serio C. (1986) Markov processes and Fourier analysis as a tool to describe and simulate daily solar irradiance. *Solar Energy*, 37, 3, 179-194.

ABSTRACT: Synthetic time series for the daily values of the clearness index (k_t) are generated for the town of Luján (Province of Buenos Aires, Argentina) for a whole year using historical monthly average values and four libraries of 10 Markov transition matrices. These arrays contain the information of the transition probabilities of the daily values of k_t of one class to another on consecutive days, for four different databases. Three of these libraries were built for this work using databases corresponding to other Argentine stations. The fourth library corresponds to a database from the northern hemisphere and was obtained from the literature. Statistical estimators to the series generated with the four libraries matrices are compared to assess which of the libraries generated the best data sets for k_t .

Key words: Synthetic Series, Markov transition matrices, clearness index.